On the Minimization of a Cost Function Defined by the NCEP Global Spectral Model with Full-physics

S. Zhang*

Geophysical Fluid Dynamics Laboratory, Princeton University

Princeton, NJ 08542, USA

X. Zou and Jon E. Ahlquist

Meteorology Department, Florida State University

Tallahassee, FL 32306, USA

Final revised version:January 2003

(submitted to *Monthly Weather Review*)

* *Corresponding author address:* Dr. S. Zhang, GFDL/NOAA, Princeton University, P.O. Box 308, Princeton, NJ 08542, USA. Email: snz@gfdl.noaa.gov

ABSTRACT

The impact of diabatic processes on the minimization of 4-dimensional variational data assimilation (4D-Var) is studied using the 1995 version of NCEP's global spectral model with and without full physics. A cost function measuring spectral errors of 6-hour forecasts to "observation" (the NCEP re-analysis data) is minimized using the L-BFGS (the limited memory quasi-Newton algorithm developed by Broyden, Fletcher, Goldfard and Shanno) for optimizing parameters and initial conditions. Minimization of the cost function constrained by an adiabatic version of the NCEP global model converges to a minimum with a significant amount of decrease in the value of the cost function. Minimization of the cost function using the diabatic model, however, fails after a few iterations due to discontinuities introduced by physical parameterizations. Examination of the convergence of the cost function in different spectral domains reveals that the large-scale flow is adjusted during the first 10 iterations, in which discontinuous diabatic parameterizations play very little role. The adjustment produced by the minimization gradually moves to relatively smaller scales between 10-20th iterations. During this transition period, discontinuities in the cost function produced by "on-off" switches in the physical parameterizations caused the cost function to stay in a shallow local minimum instead of continuously decreasing toward a deeper minimum.

Next, a mixed 4D-Var scheme is tested in which large-scale flows are first adiabatically (only with vertical diffusions and surface drag) adjusted to a sufficient level, followed by a diabatic adjustment introduced after 10 to 20 iterations. The mixed 4D-Var produced a closer fit of analysis to observations, with 38% and 41% more decrease in the values of the cost function and the norm of gradient, respectively, than the standard diabatic 4D-Var, while the CPU time is reduced by 21%. The resulting optimal initial conditions improve the short-range forecast skills of 48-hour statistics, particularly in the tropics. The detrimental effect of parameterization discontinuities on minimization was also reduced.

1 Introduction

Variational analysis, an application in atmospheric/oceanic sciences of optimal control theory (Sasaki 1958), is an approach which seeks optimal values of some or all input variables of a forecast system. Once an appropriate error function (also called a cost function, usually chosen as a quadratic form of a model forecast error) is defined, four-dimensional variational data assimilation (4D-Var) calibrates the input variables of the numerical model (the initial condition and/or model parameters, called the control variables) by minimizing the error function with respect to the control variables. This approach uses the model dynamics and physics as constraints of the optimal control problem in a four-dimensional verification space. Therefore, the solution is an optimal estimate of initial conditions and/or parameters, which always is internally consistent between the model and observations.

The traditional approach assumes that the cost function which is minimized is differentiable everywhere with respect to the control variables (Le Dimet and Talagrand 1986). This was not a problem for the early research and applications using adiabatic models (Courtier et al. 1990; Chao et al. 1992; Navon et al. 1992). Physical processes modify the local budget of mass, momentum and energy by adding sources and/or sinks and therefore play a very important role in the evolution of the atmosphere. In a discretized numerical system, many important physical processes such as turbulence and convection are incorporated into the numerical weather prediction (NWP) models through parameterization approach. The approach uses the grid-scale resolvable information to simulate the mean effect of the sub-grid scale processes in a grid-box, which is described by the derived characteristic parameters. For improving numerical forecasting by accounting for physics in data assimilation, the incremental approximation method was proposed (Courtier et al. 1994). This method uses the linearized version and adjoint of a simplified forward model to calculate the evolution of an increment and the gradient of the cost function defined by the increment to avoid demands for the adjoint of parameterized physics with its related problems. Attempts to include parameterized physics in an assimilation model and its adjoint (Zupanski 1993; Zupanski and Mesinger 1995; Zou et al. 1993b; Zou and Kuo 1996; Tsuyuki 1997; Zhu and Navon 1999; Mahfouf 1999; Janiskova et al. 1999a; Laroche et al. 2002) have revealed that when the assimilation model includes parameterized physics with "on-off" switches both the model solution and the cost function may be discontinuous. These discontinuities may render the minimization to fail. Many researchers have addressed this issue (Zupanski 1993; Xu 1996; Zou 1997; Xu and Gao

1999). Zhang et al. (2000, 2001) pointed out that a cost function involving discontinuous physical processes is piecewise differentiable and a classical adjoint integration correctly evaluates one-sided gradients at either side of a discontinuity point. In certain situations, a differentiable optimization algorithm such as the limited-memory quasi-Newton method (hereafter called L-BFGS) (Liu and Nocedal, 1989) may perform well for minimizing such a piecewise differentiable cost function since the one-sided gradients at discontinuities derived from the adjoint still provide useful information for finding search directions of minimization. However, although more expensive, a nondifferentiable optimization algorithm such as the bundle method (Lemaréchal 1978; 1989; Lemaréchal and Sagastizabal 1997) may be required in cases when the differentiable minimization algorithm fails (Zhang et al. 2000). An alternative thought on this issue is the "generalized" adjoint approach (Xu 1996; Xu and Gao 1999) which tries to account for the effect of "on-off" switches in parameterized physical processes in the tangent linear and adjoint models. Implementing these approaches in data assimilation using a realistic NWP model still is under examination.

In order to incorporate the adjustment of diabatic processes into data assimilation to extract more observational information to form initial conditions, the incremental approach (Courtier et al. 1994) was implemented for operational application (Thepaut et al. 1999; Rabier et al. 2000; Mahfouf and Rabier 2000; Klinker et al. 2000; Geleyn et al. 2001). Their experiments showed the improvements of the assimilation results in the extratropics from the incremental 4D-Var with 6-hour assimilation window over 3D-Var (3-dimensional variational data assimilation, without considering the time window). However, since the tangent linear and adjoint models used in the minimization were only approximations of the assimilating model, the incremental 4D-Var showed some problems in the tropical area and the poorer performance for a relatively long assimilation window than a 3D-Var, although the poorer performance was improved by including more elaborated physical processes (Mahfouf and Rabier 2000).

Atmospheric motion is complex with various scales influencing one another through mass, energy and momentum exchanges. Since parameterized physics attempts to simulate sub-grid processes, it is expected that the diabatic processes have the most impact on small scales. However, due to the interaction of scales, the impact of scales on the minimization in 4D-Var still is a question. For studying preconditioning issue, Courtier et al. (1994) schematically discussed some of the scale-dependence of the minimization. With the NCEP (National Centers for Environmental Prediction) global spectral model with full-physics (1995 version) and its adjoint, this study first examines the decrease of the cost functions in different wavenumber domains employing the L-BFGS algorithm

in order to gain a systematical insight about the impact of scales on the minimization. Based on the results, we tested a new 4D-Var implementation scheme to improve the minimization of a diabatic cost function.

This paper is arranged as follows: Section 2 gives a brief description of the NCEP global spectral model and its physical parameterizations, and presents the numerical results of 4D-Var experiments using the adiabatic and diabatic versions of the NCEP model and their adjoints. Section 3 applies a spectral domain decomposition of the cost function and examines the behavior of the cost function in different wavenumber domains. Section 4 tests a new 4D-Var implementation scheme to improve the performance of the L-BFGS algorithm for minimizing a cost function to include diabatic effects. The comparisons of forecasting skill for the two kinds of 4D-Var implementations are also given in this section. Conclusions and discussion are provided in section 5.

2 Performance of the L-BFGS minimization algorithm on a diabatic atmospheric model

2.1 Brief description of the NCEP global spectral model

The NCEP global spectral model consists of a set of prognostic equations, including the tendency equations of vorticity, divergence, virtual temperature, specific humidity and surface pressure (Sela 1982; 1987). The model physics include large-scale precipitation, cumulus convection, shallow convection, vertical diffusion, gravity-wave drag, surface and boundary processes. The large-scale (grid-scale) condensation that adjusts both the temperature and moisture consists of condensation in super-saturated layers and evaporation in unsaturated layers. The cumulus parameterization scheme is a simplified (one-type cloud, Pan et al. 1995) Arakawa-Schubert (1974) scheme. The shallow convection is the Betts scheme (Betts et al. 1986). A nonlocal vertical diffusion scheme is adapted (Hong and Pan 1996). The gravity-wave drag parameterization is the Pierrehumbert (1986) scheme. The surface processes include a two-layer soil model. In the related processes the Monin-Obukhov similarity theory is applied to simulate the change of the surface (air-ground interface) temperature and humidity, and to calculate exchange coefficients and turbulence fluxes in the planetary boundary layer. These physical parameterization schemes are included in both the nonlinear and the adjoint models. Long-wave and short-wave radiation are kept constant during the 6-h assimilation window.

The model dynamical core includes horizontal diffusion and time filtering in which the filter coefficient α_A is taken as $1-2\epsilon$, where ϵ is the classical Asselin filter coefficient. The model carries out the time filter and horizontal diffusion at each time-step integration. This study is conducted using the resolution with 62 waves (not including the zonal mean) in horizontal domain and 28 vertical levels.

2.2 A cost function measuring the spectral errors

In this study, in order to address the issue of the minimization in 4D-Var which includes discontinuous parameterizations we define the cost function using the spectral coefficients of the model forecasts and analyses, only taking the background term form, as

$$J(\boldsymbol{\alpha}) = \frac{1}{2}\left\langle \mathbf{W}(\mathbf{x}^{6-h\ forecast} - \mathbf{x}^{analysis}), \mathbf{x}^{6-h\ forecast} - \mathbf{x}^{analysis} \right\rangle \tag{1}$$

where $\langle , \rangle$ represents the dot product of two vectors. $\boldsymbol{\alpha}$ represents the control variable vector which may include the initial conditions $\mathbf{x}_0$ (for data assimilation) and any adjustable model parameters (for parameter estimation). In this study, $\mathbf{x}^{analysis}$ is the analysis state at 06 UTC on 1 October 1995. $\mathbf{x}^{6-h\ forecast}$ is the 6-hour model forecast state starting from the initial state at 00 UTC on 1 October 1995. The state vector $\mathbf{x}$ consists of spherical harmonic spectral coefficients of divergence ($\mathbf{D}_{m,n}$), vorticity ($\zeta_{m,n}$), virtual temperature ($\mathbf{T}_{vm,n}$), specific humidity ($\mathbf{q}_{m,n}$) and the logarithm of surface pressure ($\mathbf{P}_{m,n}$), i.e. $\mathbf{x}$=($\mathbf{D}_{m,n}$, $\zeta_{m,n}$, $\mathbf{T}_{vm,n}$, $\mathbf{q}_{m,n}$, $\mathbf{P}_{m,n}$) where m and n represent the zonal and meridional wavenumber respectively. $\mathbf{W}$ is a diagonal weighting matrix approximated by the inverse of the absolute value of maximal 6-hour differences of the variable spectral coefficients for each zonal wavenumber. In this case, $\mathbf{W}$ contains 63 (zonal wavenumbers) $\times 5$ (state variables) values. Sections 3.1 uses this spectral error cost function to investigate the decrease of the cost functions of various spatial scales on the minimization.

2.3 Jagged behavoir of a diabatic cost function leads to minimization problems

Zhang et al. (2000) discussed some aspects of the impact of discontinuities on minimization using a shallow-convection operator. In order to address minimization problems that may be caused by discontinuous physical processes in 4D-Var using a fully-parameterized NWP model, as in Zhang et al. (2001) we choose the time filtering coefficient (α_A) and the horizontal diffusion coefficient (α_{HD}) in the NCEP model to show the non-smoothness of the cost function and the corresponding minimization performance for optimizing these coefficients. Although α_A and α_{HD} are not explicitly

related to the model physical parameterizations, any change of their value can affect "on-off" switches in the parameterization schemes during the time integration. Therefore, these coefficients play the role of one-dimensional control variables. The cost function (J) defined by (1) can be evaluated for a series of values of α_A or α_{HD} by forward model integration so that the features of the J and the corresponding minimization performance can be illustrated in α_A or α_{HD} space.

Variations of the cost function J defined by (1), evaluated using the NCEP adiabatic (top) and diabatic (bottom) nonlinear forward models, with respect to α_A (left) and α_{HD} (right) are shown in Fig. 1. Values of J are computed at the interval of 0.001 for α_A and 0.001×10^{16} for α_{HD}. Comparing Fig. 1ab with Fig. 1cd, it was observed that the diabatic J appears jagged with respect to both coefficients α_A and α_{HD}. This phenomenon was explained by the piecewise differentiable nature of a diabatic cost function when the model includes discontinuous parameterizations which are turned on and off discontinuously (Zhang et al. 2001). In Fig. 1, each dark-dot indicates an iteration in minimization but since those dark-dots after 4 iterations are so close that only the first four are able to be identified. For adiabatic cases, the fourth iteration point is close to the global minimum of the cost function (panels *a* and *b*) but this is not always true in diabatic cases. For example, for the diabatic α_A case (panel *c*), the fourth iteration has $\alpha_A = 0.911$ where the value of the cost function is 762.34 (denoted by J_0 afterward) and the gradient is 2.3. The "minimum" found through forward integrating the model is located at 0.935 (signed by a star in Fig. 1c) where the value of the cost function is 762.30 (denoted by J_* afterward) and the gradient is 4×10^{-3} (panel *c*). Similar results also are found in the diabatic α_{HD} case where J_*=760.82 at α_{HD}=0.355×10^{16} (signed by a star in Fig. 1d) (the corresponding $\nabla_{\alpha_{HD}} J$=0.7) while J_0=760.97 at α_{HD}=0.364×10^{16} (the corresponding $\nabla_{\alpha_{HD}} J$=−34.8) (panel *d*).

These results indicate that when a discontinuous diabatic assimilation model is used, a differentiable optimization algorithm such as L-BFGS may not perform well in some cases. In two simple cases of one-dimensional control variable examined above, either the stationary point found by minimization is not a global minimum or the minimization is locked at some discontinuity point rather than approaching a stationary point. The latter problem is addressed in this study. Even though in this simple case of one-dimensional control variable the difference of J_* and J_0 is small, it clearly tells that discontinuities in diabatic model may hinder the minimization to approach a deeper minimum. The further examination in next section will find that in a multi-dimensional minimization problem such as data assimilation, this kind of stagnation of minimization may cause a significant error for initial conditions which are optimized.

2.4 4D-Var experiments using adiabatic and diabatic atmospheric models

With the adjoints (Zhang et al. 2001) of the adiabatic and diabatic versions of the NCEP global spectral model, we carry out numerical experiments for 4-dimensional variational data assimilation employing the L-BFGS algorithm. Hereafter, the 4D-Var experiment using the diabatic model and its adjoint is called "diabatic 4D-Var" and the resulting optimal initial conditions are called "diabatic optimal" initial conditions while the experiment using the adiabatic model and its adjoint is called "adiabatic 4D-Var" and the resulting optimal initial conditions are called "adiabatic optimal" initial conditions.

Starting from the first guess, the NCEP re-analysis data at 00 UTC 1 October 1995, the adiabatic and diabatic 4D-Var experiments adjust the initial conditions by minimizing the defined cost function by (1). The variations of the cost function J (top) and the norm of gradient ($\|\nabla J\|$) (bottom) with iteration numbers are exhibited in Fig. 2. Figure 2 shows that at the initial stage of the minimization, the diabatic cost function is smaller than the adiabatic one since a diabatic model simulates the atmospheric state more realistically. During the first 10 iterations, both adiabatic and diabatic cost functions and the norms of their gradient are rapidly reduced by 62% and 90% (for the adiabatic J and $\|\nabla J\|$), and 54% and 83% (for the diabatic J and $\|\nabla J\|$). However, the rate of decrease of the cost function of the diabatic model slows after 10 iterations. After 45 iterations, the diabatic J and $\|\nabla J\|$ are even unchanged, which suggests that the minimization be locked. This leads to an opposite difference of the adiabatic and diabatic J's at the end of minimization, iteration 60, as at the beginning, i.e., 60 iterations of minimization reduce the adiabatic J and $\|\nabla J\|$ by 88% (from 919 to 108) and 95% (from 284 to 13), but only 66% (from 762 to 263) and 88% (from 293 to 34) for the diabatic J and $\|\nabla J\|$.

The reason which causes the insufficient decrease of the diabatic cost function is owing to discontinuities in the diabatic model (Zhang et al. 2000). However, for a multi-dimensional proplem, it is inconvenient to graphically demonstrate that the minimization is locked by discontinuities. The study of using the idealized one-dimensional stepping function (Zhang et al. 2000) found that when the minimization is locked in a discontinuity point, the adjusted variable usually goes back and forth around the discontinuity point as iteration proceeds. Therefore, for multi-dimensional control variable situations, we can check the adjustment of control variables at neighboring iterations to examine the performance of minimization. For the diabatic 4D-Var experiment above, around iteration 50, the adjustment of control variables in minimization [the adjustment of a variable

at iteration-n is defined as the difference between the values of the variable at iteration-n and iteration-(n-1)] is dominated by two centers at tropics over the west Atlantic and Indian Ocean. The diabatic adjustment center over the west Atlantic of the specific humidity at the 5th model level ($\sigma = 0.931$) at iteration 49 (top), 50 (middle) and 51 (bottom) is shown in the right panel of Fig. 3. The corresponding adiabatic adjustment is shown in the left panel of Fig. 3. Fig. 3 shows that the adjustment of control variables in the diabatic minimization alternatively switches the sign with the same patterns as iteration is proceeding, which leads to the stagnation of the diabatic J, while the adiabatic minimization continuously adjusts the relatively small scale distributions of the variables for reducing the adiabatic J. The behavoir of the diabatic minimization in which the adjusted variable goes back and forth around a state reveals a typical feature of the locked minimization at discontinuities.

These examinations for one- (section 2.3) and multi-dimensional minimization problems show that the discontinuous parameterizations in the diabatic model can cause the cost function to be locked at a local minimum instead of continuously decreasing toward a deeper minimum. This is consistent with the results using only the shallow-convection parameterization (Zhang et al. 2000). In that simple case, a nondifferentiable optimization algorithm, the bundle method (Lemaréchal 1978; 1989; 1997), is used to gain an improved minimization (Zhang et al. 2000). However, for a realistic atmospheric model, a nondifferentiable optimization algorithm is very expensive. In addition, although the bundle method has been tested successfully for certain individual parameterizations, its practical application for realistic models is under examination. On the other hand, even though the "generalized" adjoint (Xu 1996) is a promising thought on the minimization for discontinuous functions and was tested (Xu and Gao 1999) using the Lorenz-63 model, unfortunately, implementing the approach in data assimilation using a primitive equation model including complicated physical parameterization schemes seems difficult and requires further study. Therefore, improving the performance of the L-BFGS algorithm on a diabatic assimilation model is quite significant for both research and operational application of 4D-Var.

2.5 Verification of forecasting errors using the adiabatic and diabatic 4D-Var results as initial conditions

In order to examine the impact of the adiabatic and diabatic "optimal" initial conditions on forecast skill, the adiabatic and diabatic models are advanced to 48 hours starting from these "optimal" initial conditions that are rendered at the end of the minimization (the 60th iteration). For each

adiabatic or diabatic model, 3 runs are conducted: the control run, the optimal run and the cross run. The 'cross run' is an adiabatic run using diabatic optimal initial conditions, called 'adiabatic cross run' or a diabatic run using adiabatic optimal initial conditions, called 'diabatic cross run', while the control/optimal run uses the first guess/optimal initial conditions. The root mean square (RMS) errors of wind speed (u, v), virtual temperature (T_v) and specific humidity (q) 48-hour forecasts in 6 forecasting runs are listed in Table 1, in which statisics are conducted over the domain of 30°S to 30°N and the bottom 10 levels in order to address the impact of diabatic processes.

Table 1 shows that over this verification domain two diabatic forecasts, the diabatic optimal run and diabatic cross run, have the smallest RMS errors; the diabatic optimal run generally has a slightly smaller total RMS error than the diabatic cross run. Particularly concerning, we examined the vertical distribution of the RMS errors of the specific humidity as shown in Fig. 4. Figure 4 shows that the diabatic optimal run has smaller RMS errors of the specific humidity than the diabatic cross run at most levels. Obviously, a diabatic 4D-Var procedure including moist processes should be able to adjust the moisture field more reasonably. Considering that the value of the diabatic cost function at the 60th iteration is more than twice as large as the adiabatic one (263 to 108), a diabatic 4D-Var with a sufficiently-decreased cost function should produce better initial conditions for a diabatic model which simulates a more realistic evolution of the atmosphere. The question is how can we obtain a sufficient decrease of the cost function of a diabatic model using the L-BFGS algorithm?

In the following sections, a mixed (adiabatic/diabatic) 4D-Var scheme is designed to solve this problem, leading to a faster convergence rate for the diabatic minimization at a reduced computational cost.

3 Decrease of the cost functions of different scale flows in the L-BFGS minimization

3.1 Partition of the cost function in spectral domains

The inner product definition of the cost function in (1) can be expanded as a weighted sum of squared errors of spectral components of the model state variables as

$$J(\boldsymbol{\alpha}) = \frac{1}{2} \quad \{ \quad \sum_{m=0}^{M} \sum_{n=m}^{M} w_D(m) \sum_{k=1}^{K} [(e_D)_{m,n}]_k^2 + \sum_{m=0}^{M} \sum_{n=m}^{M} w_\zeta(m) \sum_{k=1}^{K} [(e_\zeta)_{m,n}]_k^2$$

$$+ \sum_{m=0}^{M}\sum_{n=m}^{M} w_T(m)\sum_{k=1}^{K}[(e_{T_v})_{m,n})]_k^2 + \sum_{m=0}^{M}\sum_{n=m}^{M} w_q(m)\sum_{k=1}^{K}[(e_q)_{m,n})]_k^2$$
$$+ \sum_{m=0}^{M}\sum_{n=m}^{M} w_P(m)\,[(e_P)_{m,n})]^2\Big\} \qquad (2)$$

where $(e_x)_{m,n}$ represents the (m,n)-th component of spherical harmonic spectral errors (here the subscript x indicates the physical variable) and $w_x(m)$ is a weighting constant for a given physical variable for the m-th zonal Fourier spectrum. For example, $(e_D)_{m,n}$ is the (m,n)-th component of the spectral error of divergence and the corresponding weighting is w_D(m). $\sum_{m=0}^{M}\sum_{n=m}^{M}$ describes a sum over the spectral component indices and $\sum_{k=1}^{K}$ describes a sum over the vertical level indices and M and K represent respectively the total zonal wavenumber (in this case, M=62) and the total number of vertical levels (K=28).

The NCEP global spectral model uses a triangular trancation scheme to carry out the spherical-harmonic transformation. For example, the 0-th zonal Fourier spectrum represents the zonal mean and the meridional oscillation of the zonal mean is represented by a Legendre sum from 0 to M. However, for the m-th zonal Fourier spectrum, its meridional oscillation is only transformed to a Legendre sum from m to M. Therefore, a two-dimensional array which holds the spectral coefficients is arranged as a lower triangle form like (0,0), (1,0), (2,0),...,(M,0), (1,1), (2,1),..., (M,1), (2,2), (3,2),... If $I_{m,n}$ is used to index these spectral components, the total cost function as a sum of the contributions of different spectral domain is

$$J(\boldsymbol{\alpha}) = \sum_{I_{m,n}=1}^{(M+1)\times(M+2)/2} J_{I_{m,n}} = J_{L_w} + J_{M_w} + J_{S_w}. \qquad (3)$$

$J_{I_{m,n}}$ represents the contribution of a spectral component to $J(\boldsymbol{\alpha})$ and the subscripts $_{L_w}$, $_{M_w}$ and $_{S_w}$ represent a re-grouping of the spectral domains in which the total cost function (J) can be partitioned to study the contribution of various scale flows for J. Then,

$$J_{I_{m,n}} = \frac{1}{2}\ \{\ w_D(m)\sum_{k=1}^{K}[(e_D)_{m,n}]_k^2 + w_\zeta(m)\sum_{k=1}^{K}[(e_\zeta)_{m,n}]_k^2$$
$$+\ w_T(m)\sum_{k=1}^{K}[(e_{T_v})_{m,n})]_k^2 + w_q(m)\sum_{k=1}^{K}[(e_q)_{m,n})]_k^2$$
$$+\ w_P(m)\,[(e_P)_{m,n})]^2\Big\}, \qquad (4)$$

and

$$J_{L_w} = \sum_{m=0}^{M_1}\sum_{n=m}^{M_1} J_{I_{m,n}}$$

$$
\begin{aligned}
J_{M_w} &= \sum_{m=0}^{M_2} \sum_{n=M_1}^{M_2} J_{I_{m,n}} \\
J_{S_w} &= \sum_{m=0}^{M} \sum_{n=M_2}^{M} J_{I_{m,n}}
\end{aligned} \tag{5}
$$

If M_1 and M_2 are appropriately chosen, J_{L_w}, J_{M_w} and J_{S_w} represent the contribution of the "Long-wave," "Middle-wave" and "Short-wave" flows in the evolution of the atmosphere. The decrease of J_{L_w}, J_{M_w} and J_{S_w} can be examined to identify the adjustment of different scales at different minimization stages.

Figure 5 exhibits the spectrum of the virtual temperature at the 9th model level (from the ground, $\sigma = 0.778$) with $I_{m,n}$ at 00 UTC on 1 October 1995 where the spectrum for $I_{m,n} = 0$ to 252 is magnified in the top-right window. The main power of the spectrum concentrates on the zonal waves 0-9. The magnified part shows the main power of the spectrum for each zonal wave concentrates in the first 10 meridional waves. Here we attempt to separate roughly the relative scales of the atmospheric flows, instead of proposing an elaborate cut-off of wavenumbers for decomposition. For example, from Fig. 5, we pick $M_1 = 10$ and $M_2 = 20$ to form J_{L_w}, J_{M_w} and J_{S_w} to give three wavenumber groups: 0-10, 11-20, 21-62.

Figure 6 presents the distribution of the virtual temperature at the 9th model level at 00 UTC on 1 October 1995 and its decomposition in three spectral domains corresponding to "L_w," "M_w" and "S_w" described above. Figure 6 shows that these spectral domains decompose the original virtual temperature field (panel *a*) into the different scales (panel *b*, *c* and *d*). This decomposition will be used to examine the decrease of the cost function in different spectral domains in the next section.

3.2 Decrease of the cost function in different spectral domains with the L-BFGS minimization

In order to study the adjustment of different scales at different stages of the minimization, the cost function at each iteration in the top panel of Fig. 2 is decomposed by the method developed in the last section. The variations of these cost functions in three wavenumber domains, J_{L_w}, J_{M_w} and J_{S_w}, with the iteration number are shown in Fig. 7, in which the bottom part (the value of the cost functions less than 100) is magnified in the right-top window. From Fig. 7, we found that at the beginning stage (around the first 8 iterations) of minimization, J_{L_w} decreases at the fastest rate in both adiabatic (reduced by 78%) and diabatic (reduced by 77%) cases and all diabatic cost

functions J_{L_w}, J_{M_w} and J_{S_w} have almost the same rate of decrease as the corresponding adiabatic functions. This means that at this stage the minimization mainly adjusts larger scale flows and discontinuities in parameterizations are not important.

As iteration proceeds, the adjustment of the relatively small scales increases and discontinuities in parameterizations have an increasing impact on the minimization. This leads to a slower rate of decrease of the diabatic J_{L_w} during the minimization period from iteration 9 to 30. During the transition period, the decrease of the diabatic J_{L_w} and J_{M_w} is impacted by discontinuities due to the feedback on the large-scale flows from small scales. After 30 iterations, the diabatic cost functions in all three wavenumber domains (J_{L_w}, J_{M_w} and J_{S_w}) are almost unchanged while in the adiabatic case, J_{S_w} has the fastest decrease rate. This means that after 30 iterations, the minimization adjustment focuses on relatively small scales. The evolution of the minimization can be further revealed by examining the adjustment magnitude of state variables in three wavenumber domains through one iteration at different minimization stages. For example, Fig. 8 exhibits the change of the "L_w" virtual temperature at the 9th model level at iteration 5 (panel a), iteration 20 (panel b) and iteration 50 (panel c) in the adiabatic case. Fig. 8 shows that the magnitude of large scale adjustment at iteration 5 (panel a) is 5 times larger than the adjustment at iteration 20 and 10 times larger than the adjustment at iteration 50. This means that the large scale adjustment made by minimization around iteration 50 is very small. The corresponding adjustment magnitudes of "M_w" and "S_w" are shown in Fig. 9 and Fig. 10. From Fig. 9, the "M_w" change of the virtual temperature at iteration 5, iteration 20 and iteration 50 increases compared with "L_w", especially at iterations 20 and 50. Figures 10 shows that the minimization adjustment gradually moves to relatively small scales after the transition period.

For the diabatic situation, the adjustments of the virtual temperature of "L_w," "M_w" and "S_w" at iteration 5 have the similar features as the adiabatic adjustments (not shown). The diabatic adjustments of the virtual temperature at the 9th model level of "L_w," "M_w" and "S_w" at iteration 20 and 50 are shown in Fig. 11 and Fig. 12. From Fig. 11, we find that, at iteration 20, the diabatic adjustment starts to emphasize diabatic processes such as convections over tropics. At iteration 50 (Fig. 12), the "S_w" diabatic adjustment is totally dominated by only two strong centers at tropics. Apparently, these two strong adjustment centers are produced by discontinuous physical processes. Due to feedback, the "M_w" and "L_w" diabatic adjustments at iteration 50 also only exist over the regions where two "S_w" adjustment centers come out. These features of diabatic adjustments are consistent with the nature of the locked minimization by discontinuities at this stage (section 2.4,

Fig. 3).

It is worth mentioning that the results above are based on the NCEP reanalysis data as "observations." When real observations are used, additional considerations may be required. For example, if only the small scale features are "observed," it is possible that small scale features in the real observations may have a considerable impact at the beginning of minimization. This was demonstrated by Thépaut and Courtier (1991) who conducted a twin experiment study on this issue. The other point worth mentioning is that for a more realistic assimilation experiment, the scale impact from the background and gravity wave penalty terms should be considered.

From the above examination and analysis about the impact of scales on minimization, in the next section, a mixed 4D-Var procedure which combines the adiabatic and diabatic models and their adjoints is designed to obtain a better decrease of the diabatic cost function in the L-BFGS minimization at a reduced computational cost.

4 A mixed 4D-Var experiment scheme and its numerical results

4.1 Experiment design

The analysis in section 3.2 showed that the L-BFGS minimization first adjusts large-scale flows and then relatively small scales as iteration proceeds. The large scale adjustment proceeds rapidly through the first 10 iterations. Next, the transition period lasts approximately from 10 to 20 iterations. During the transition period, discontinuities in parameterizations may cause a diabatic cost function to stagnate instead of continuously decreasing as iteration continues. A mixed 4D-Var scheme tries to skip this critical period of transition in which discontinuities in parameterizations have the most impact on the minimization by employing an adiabatic model and adjoint. Then, the diabatic model and its adjoint are invoked to adjust the relatively small scales. The mixed 4D-Var scheme is expected to adjust the relatively smaller scale flows better after the adiabatic adjustment allows the minimization to continue through the transition period. In addition, the mixed scheme will be cheaper than a pure diabatic scheme since the adiabatic computation at the early stage costs much less than the diabatic version. Since no diabatic processes are involved in the first adiabatic phase, a trade-off of this kind of mixed 4D-Var scheme is that no observational information related to these processes is assimilated in the minimization period.

4.2 Numerical results

The study tests three switch schemes from the adiabatic adjustment to the diabatic adjustment: the adiabatic model and its adjoint are replaced by the diabatic ones at iteration 10, 15 and 20 (called mix10, mix15 and mix20 hereafter). Figure 13 exhibits the variations of the cost function (top) and the norm of gradient (bottom) with iteration number for mix10, mix15 and mix20, in which the original adiabatic and diabatic experiments (from Fig. 2) are plotted using the thin-dotted line (adiabatic) and the thin-solid (diabatic) line as the reference. Both mix15 and mix20 improve the minimization of the diabatic cost function. However, mix15 produces the most decrease of the diabatic cost function by 60 iterations in these diabatic 4D-Var experiments while mix10 produces the same decrease of the diabatic cost function as the old scheme does after around 25 iterations. In addition, experiments using 12-hour assimilation window also showed that mix15 made the most reduction for the cost function (from 431 to 316) and the norm of the gradient (from 51 to 30) around 40 iterations (not shown). The experiments suggest that the 15-iteration adiabatic adjustment allow the minimization process to have gone through some critical period of transition so that discontinuities in diabatic model have less impact on the minimization afterward.

Comparing to the original diabatic 4D-Var experiment, mix15 (6-hour assimilation case) reduces the cost function (J) and the norm of gradient ($\|\nabla J\|$) by 38% and 41% more and mix20 reduces by 29% and 35% more respectively. The vertical distributions of the RMS errors of u, v, T_v and q at the end of the assimilation window (6-hour forecasts) for the original scheme, mix15 and mix20 are presented in Fig. 14 which shows that the RMS errors of all variables are reduced in the new mixed scheme at almost all levels. In order to examine the improvement of forecast skill resulting from the new assimilation scheme, the RMS errors of the diabatic model forecasts using the optimal initial conditions derived by the new assimilation scheme as a function of forecast lead time are listed in Table 2, in which statistics are conducted over $30^\circ S$ to $30^\circ N$. Table 2 shows that the RMS errors of forecasts of all u, v, T_v and q have been reduced throughout 48 hours.

Computation times for 5 experiments: adiabatic, diabatic, mix10, mix15 and mix20, are given in Fig. 15. Mix10 saves 10% on CPU (central processing unit) time and mix15 and mix20 save 15% and 21% CPU time respectively.

The numerical experiments above showed that the mixed minimization scheme is able to reduce the diabatic cost function more at a cheaper computational cost.

5 Summary and Discussion

The impact of diabatic processes on 4-dimensional variational data assimilation (4D-Var) was studied using the 1995 version of NCEP's global spectral model with and without full physics. First, the NCEP re-analysis data on 1 October 1995 were used to carry out adiabatic and diabatic 4D-Var experiments by employing the limited memory quasi-Newton algorithm (L-BFGS) for optimizing parameters and initial conditions. The cost function minimized measured spectral errors of 6-hour forecasts to "observations" (the NCEP's reanalysis data). During minimization, the cost function of the diabatic model was locked in a shallow local minimum due to discontinuities in parameterizations while the minimization of the cost function of the adiabatic model converged to a minimum with a significant amount of decrease of the cost function.

In order to improve the performance of the L-BFGS algorithm in the diabatic model, the decrease of the cost function for different spectral domains was examined. At the beginning stage, say, before the tenth iteration, the minimization process mainly adjusts the large scale flows, in which discontinuous diabatic parameterizations play little role. As minimization continues, the adjustment gradually moves to relatively small scales between 10-20 iterations. During this transition period, discontinuities in parameterizations of model physics may cause the cost function to stay in a shallow local minimum instead of continuously decreasing toward a deeper minimum.

Next, a mixed 4D-Var implementation scheme was designed. The scheme first uses the adiabatic model and its adjoint to adjust large-scale flows. After 10 to 20 iterations, the diabatic model and its adjoint are used to minimize the cost function for adjusting relatively small scales. Results show that this mixed scheme reduces the diabatic cost function and the norm of gradient up to 38% and 41% more respectively for 60 iterations than the old scheme. Further, it reduces CPU time up to 21% on the NCAR Cray cluster. The resulting optimal initial conditions from the new scheme improve the short-range forecast skill with 48-hour statistics.

This study suggests that one may solve a diabatic 4D-Var problem using an economical differentiable optimization algorithm, with the cheaper two-step minimization scheme. In the mixed 4D-Var scheme, the best switch from adiabatic adjustment to diabatic adjustment for decreasing the diabatic cost function can be chosen between 10 to 20 iterations. Considering the trade-off of the diabatic adjustment and the computational cost, the experiments in this study suggests a best switch around 15-iteration. It is worth mentioning that the mixed 4D-Var scheme only lessens the

effect of parameterization discontinuities on the diabatic minimization but not entirely gets free of the problem. In the future, implementing and testing the bundle method (Lemaréchal 1978; 1989; Lemaréchal and Sagastizabal 1997) and the "generalized" adjoint approach (Xu 1996; Xu and Gao 1999) in a realistic NWP model is necessary.

ACKNOWLEDGEMENT

The research is supported by NSF grant ATM-9812729 and NOAA grant NA77WA0571. Authors would like to thank Jeff Anderson, Tony Rosati and Matt Harrison for their comments on earlier versions of this manuscript. Thanks go to Drs. E. Kalnay and J. Sela for their persistent support and encouragement. Computations were carried out on the NCAR Cray cluster and the data were taken from the NCAR data library with the SCD/NCAR grand project-35111115.

REFERENCES

Arakawa, A. and W. H. Schubert, 1974: Interaction of a cumulus ensemble with the large-scale environment, Part I. *J. Atmos. Sci.,* **31**, 674–701.

Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.,* **112**, 677–691.

Buizza, R., 1994: Sensitivity of optimal unstable structure. *Quart. J. Roy. Meteor. Soc.,* **120**, 429–451.

Courtier, P. and Talagrand O., 1987: Variational assimilation of meteorological observations with the adjoint equation - Part I. Numerical results. *Quart. J. Roy. Meteor. Soc.,* **113**, 1329–1347.

Courtier, P. and J. N. Thepaut and O. Talagrand, 1990: 4-dimensional data assimilation using the adjoint of a primitive equation model. WMO. International Symposium on Assimilation of observations in meteorology and oceanography, Clermont-Ferrand, France, 337-340.

Courtier, P. and J.-N. Thepaut and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120**, 1661–1673.

Fillion, L. and R. Errico, 1997: Variational assimilation of precipitation data using moist convective parameterization schemes: A 1D-Var study. *Mon. Wea. Rev.,* **125**, 2917–2942.

Geleyn, F., D. Banciu, M. Bellus, R. El Khatib, P. Moll, P. Saez and J.-N., 2001: The operational 4D-Var data assimilation system of Meteo-France: Specific characteristics and behaviour in the special case of the 99 Xmas storms over France. In *Proceedings of 18th Conference on Weather Analysis and Forecasting and 14th Conference on Numerical Weather Prediction*, 30 July - 2 August 2001, Fort Lauderdale, Florida, USA

Hong, S. Y. and H. L. Pan, 1996: Nonlocal boundary-layer vertical diffusion in a medium-range model. *Mon. Wea. Rev.,* **124**, 2322–2339.

Janiskova, M., J.-N. Thepaut and J.-F. Geleyn, 1999a: Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Mon. Wea. Rev.,* **127**, 26–45.

Janiskova, M., F. Veerse, J.-N. Thepaut, G. Desroziers and B. 1999b: Impact of a simplified physical package in 4D-Var analyses of FASTEX situations. *Quart. J. Roy. Meteor. Soc.,* **125A**, 2465–2485.

Klinker, E., F. Rabier, G. Kelly, J.-F. Mahfouf, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. *Quart. J. Roy. Meteor. Soc.,* **126**, 1191–1215.

Kuo, Y.-H., X. Zou, and Y.-R. Guo, 1995: Variational assimilation of precipitable water using a nonhydrostatic mesoscale adjoint model. Part I: Moisture retrieval and sensitivity experiments. *Mon. Wea. Rev.,* **124**, 122–147.

Laroche, S., M. Tanguay and Y. Delage, 2002: On the linearization of a simplified planetary boundary layer parameterization. *Mon. Wea. Rev.,* **130**, 2074–2087.

Le Dimet, F. X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus,* **38A**, 97–110.

Lemaréchal, C., 1978: Nonsmooth optimization and descent methods. International Institute for Appl. Sys. Analysis, Laxenburg, Austria, 25pp

Lemaréchal, C., 1989: Nondifferentiable optimization. *Optimization, Handbooks in OR&MS*, **1**, G. L. Nemhauser et al. (eds.), Elsevier Science Publishers, 529–572.

Lemaréchal, C. and C. Sagastizabal 1997: Variable metric bundle methods: From conceptual to implementable forms. *Mathematical Programming*, **76**, 393–410.

Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Mathematical Programming,* **45**, 503–528.

Mahfouf, J.-F., 1999: Influence of physical processes on the tangent-linear approximation. *Tellus,* **51A**, 147–166.

Mahfouf, J.-F., and F. Rabier 2000: The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. *Quart. J. Roy. Meteor. Soc.,* **126**, 1171–1190.

Navon, I. M. and D. M. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.,* **115**, 1479–1502.

Navon, I. M., X. Zou, J. Derber, and J. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.,* **120**, 1433–1446.

Pan, H. L. and W. S. Wu, 1995: Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. *Techical Report for NMC/NOAA/NWS*, No. **409**.

Pierrehumbert, R. T., 1986: An essay on the parameterization of orographic gravity wave drag. *GFDI/NOAA*, Princeton University, Princeton, NJ08542.

Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.,* **126**, 1143–1170.

Rabier, F., J.-N. Thepaut and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.,* **124**, 1861–1887.

Tanguay, M., P. Bartello and P. Courtier, 1991: Four-dimensional data assimilation with a wide range of scales. *Tellus,* **48A**, 96–121.

Sela, J., 1987: The new T80 NMC operational spectral model. *Preprint, the Eighth conference on Numerical Weather Prediction, Baltimore, Maryland, February 22–26,1988.*

Thepaut, J. N. and P. Courtier, 1991: 4-dimensional data assimilation using the adjoint of a multilevel primitive equation model. *Quart. J. Roy. Meteor. Soc.,* **117**, 1225–1254.

Thepaut, J.-N., M. Janiskova, P. Gauthier, G. Desroziers, G. Hello, B. Pouponneau and F. Veerse, 1999: Towards an operational 4D-Var assimilation system in Meteo-France. In *Proceedings of the Third WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography*, 7-11 June 1999, Quebec City, Canada.

Tsuyuki, T., 1997: Variational data assimilation in the tropics using precipitation data. Part III: Assimilation of SSM/I precipitation rates. *Mon. Wea. Rev.,* **125**, 1447–1464.

Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan,* **36**, 738–742.

Verlinde, J. and W. R. Cotton, 1993: Fitting microphysical observations of nonsteady convective clouds to a numerical model: An application of the adjoint technique of data assimilation to a kinematic model. *Mon. Wea. Rev.,* **121**, 2776–2793.

Xu, Q., 1996: Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.,* **53**, 1123–1142.

Xu, Q., and J. Gao, 1999: Generalized adjoint for physical processes with parameterized discontinuities - Part VI: Minimization problems in multi-dimensional space. *J. Atmos. Sci.,* **56**,

994-1002.

Xiao Q., and X. Zou, and Y.-H. Kuo, 1999: Incorporating the SSM/I derived precipitable water vapor and rain rate into a numerical model: A Case study for ERICA IOP-4 cyclone. *Mon. Wea. Rev.,* **128**, 87–108.

Zhang, S., X. Zou, J. Ahlquist, I. M. Navon and J. G. Sela, 2000: Use of differentiable and nondifferentiable optimization algorithms for variational data assimilation with discontinuous cost functions. *Mon. Wea. Rev.,* **128**, 4031–4044.

Zhang, S., X. Zou, J. Ahlquist, 2001: Examination of numerical results from tangent linear and adjoint of discontinuous nonlinear models. *Mon. Wea. Rev.,* **129**, 2791-2804.

Zhu, Yanqiu and I.M. Navon, 1999: Impact of key parameters estimation on the performance of the FSU spectral model using the full physics adjoint. *Mon. Wea. Rev.,* **127**, 1497–1517.

Zou, X., I. M. Navon, M. Berger, Paul K. H. Phua, T. Schlick, and F. X. LeDimet, 1993a: Numerical experience with limited-memory quasi-Newton and truncated-Newton methods. *SIAM Journal on Optimization*, **3**, 582–608.

Zou, X., Navon, I. M. and Sela, J. G. 1993b. Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus,* **45A**, 370–387.

Zou, X., 1997: Tangent linear and adjoint of "on-off" processes and their feasibility for use in 4-dimensional variational data assimilation. *Tellus,* **49A**, 3–31.

Zou, X., and Y.-H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.,* **124**, 2859–2882.

Zupanski, D., 1993: The effects of discontinuities in the Betts-Miller cumulus convection scheme on four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Tellus,* **45A**, 511–524.

Zupanski, D. and F. Mesinger, 1995: Four-dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.,* **123**, 1112–1127.

TABLES

Table 1: RMS errors of 48-hour forecasts over 30°S – 30°N and bottom 10 levels in 6 forecasting cases

runs	RMS errors			
	u(m/s)	v(m/s)	T_v (K)	q (0.1 k/kg)
adiabatic control	3.84	4.00	3.52	6.53
diabatic control	3.92	3.85	2.92	5.64
adiabatic optimal	3.87	3.71	3.47	6.40
diabatic cross	3.62	3.31	2.85	5.57
adiabatic cross	3.84	3.69	3.53	6.42
diabatic optimal	3.61	3.30	2.83	5.54

Table 2: RMS errors over $30°S$ to $30°N$ and bottom 10 levels in the standard (old) 4D-Var scheme, mix15 and mix20 respect with to forecast leading time

forecast time (hour)	RMS errors											
	u (m/s)			v (m/s)			$T_v(K)$			q (0.1 g/kg)		
	old	mix15	mix20	old	mix15	mix20	old	mix15	mix20	old	mix15	mix20
06	0.90	0.81	0.83	0.79	0.68	0.71	1.07	1.00	1.02	1.44	1.08	1.18
12	2.31	2.25	2.26	2.58	2.51	2.54	1.77	1.71	1.73	3.07	2.85	2.93
18	2.60	2.56	2.57	2.72	2.64	2.67	2.08	2.02	2.03	3.69	3.44	3.55
24	2.76	2.71	2.73	2.83	2.75	2.77	2.29	2.24	2.25	4.14	4.00	4.05
30	2.97	2.92	2.94	2.95	2.89	2.91	2.38	2.33	2.35	4.45	4.33	4.38
36	3.12	3.08	3.10	3.02	2.97	3.00	2.48	2.43	2.44	4.69	4.58	4.61
42	3.41	3.37	3.38	3.16	3.11	3.13	2.80	2.76	2.78	5.05	4.95	4.98
48	3.61	3.57	3.59	3.30	3.27	3.28	2.83	2.80	2.81	5.54	5.43	5.48

FIGURE CAPTIONS

Fig. 1 Variations of the cost function (J), defined as a weighted sum of the squared spectral errors of 6-hour forecasts (see (1)), with respect to the time filtering coefficient, α_A (left), and the horizontal diffusion coefficient, α_{HD} (right), using the adiabatic (top) and diabatic (bottom) version of the NCEP model: (a) adiabatic model, α_A, (b) adiabatic model, α_{HD}, (c) diabatic model, α_A and (d) diabatic model, α_{HD}. The cost function is evaluated by a 0.001 interval of α_A and a 0.001×10^{16} interval of α_{HD}. A dark-dot and a number beside it indicate the minimization position and the corresponding iteration number, for solving an optimal parameter value. "Observations" are the NCEP re-analyses at 00 and 06 UTC on 1 October 1995.

Fig. 2 Variations of the cost function (J) (top) and the norm of gradient ($\|\nabla J\|$) (bottom) using the NCEP adiabatic (dashed) and diabatic (solid) global spectral models, with respect to iteration number in the minimization, for optimizing initial conditions.

Fig. 3 The change of the specific humidity at the 5th model level ($\sigma = 0.931$) at iteration 49 and 48 (top), 50 and 49 (middle), 51 and 50 (bottom) in adiabatic (left) and diabatic (right) 4D-Var minimization. The interval of contours is 0.05×10^{-3} kg kg^{-1} and dashed contours are negative and solid contours are positive.

Fig. 4 Vertical distributions of RMS errors of specific humidity of 48-hour forecasts over 30°S to 30°N in three diabatic forecast cases: control run (using the NCEP re-analysis as initial conditions: dotted line), diabatic cross run (using the adiabatic optimal initial conditions: thick-dashed line) and diabatic optimal run (using optimal initial conditions solved by a diabatic 4D-Var experiment: thick-solid line) over the global domain.

Fig. 5 Spectrum of virtual temperature on the 9th model level (from the ground, $\sigma = 0.7508$) with respect to the spectral component index. The first 252 components are magnified in the top-right window.

Fig. 6 Distributions of a) the original temperature, b) the temperature of "Large-waves" (wavenumbers less than 10), c) the temperature of "Middle-waves" (wavenumbers between 11 and 20) and d) the temperature of "Short-waves" (wavenumbers greater than 20) at the 9th model level, from the NCEP re-analysis at 00 UTC on 1 October 1995. The interval of contours is 4 K for (a), (b) and (c), 2 K for (d).

Fig. 7 Variations of the cost functions in different wavenumber ranges: "Large-waves" (J_{L_w}), "Middle-waves" (J_{M_w}) and "Short-waves" (J_{S_w}) with iteration number in adiabatic and diabatic cases for optimizing initial conditions. The bottom part (the cost function less than 100) is magnified in the top-right window.

Fig. 8 Changes of the "L_w" virtual temperature at the 9th model level ($\sigma = 0.778$) at (a) iteration 5, (b) iteration 20 and (c) iteration 50 in an adiabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c).

Fig. 9 Changes of the "M_w" virtual temperature at the 9th model level ($\sigma = 0.778$) at (a) iteration 5, (b) iteration 20 and (c) iteration 50 in an adiabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c).

Fig. 10 Changes of the "S_w" virtual temperature at the 9th model level ($\sigma = 0.778$) at (a) iteration 5, (b) iteration 20 and (c) iteration 50 in an adiabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c).

Fig. 11 Changes of the (a) "L_w," (b) "M_w" and (c) "S_w" virtual temperature at the 9th model level ($\sigma = 0.778$) at iteration 20 in a diabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c). The dashed contours are negative and solid contours are positive.

Fig. 12 Changes of the (a) "L_w," (b) "M_w" and (c) "S_w" virtual temperature at the 9th model level ($\sigma = 0.778$) at iteration 50 in a diabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c). The dashed contours are negative and solid contours are positive.

Fig. 13 Variations of the cost function (J) (top) and the norm of gradient (bottom) ($\|\nabla J\|$) with iteration number in the new mixed 4D-Var schemes (mix10, mix15 and mix20). The corresponding J and $\|\nabla J\|$ in the old scheme (from Fig. 2) are plotted as reference.

Fig. 14 Vertical distributions of the RMS errors of u, v, T_v and q in mix15 (thick-solid), mix20 (dashed) and old (thin-solid) schemes at the end of assimilation window.

Fig. 15 CPU time to conduct 60 iterations for solving the optimal initial conditions in 5 cases.

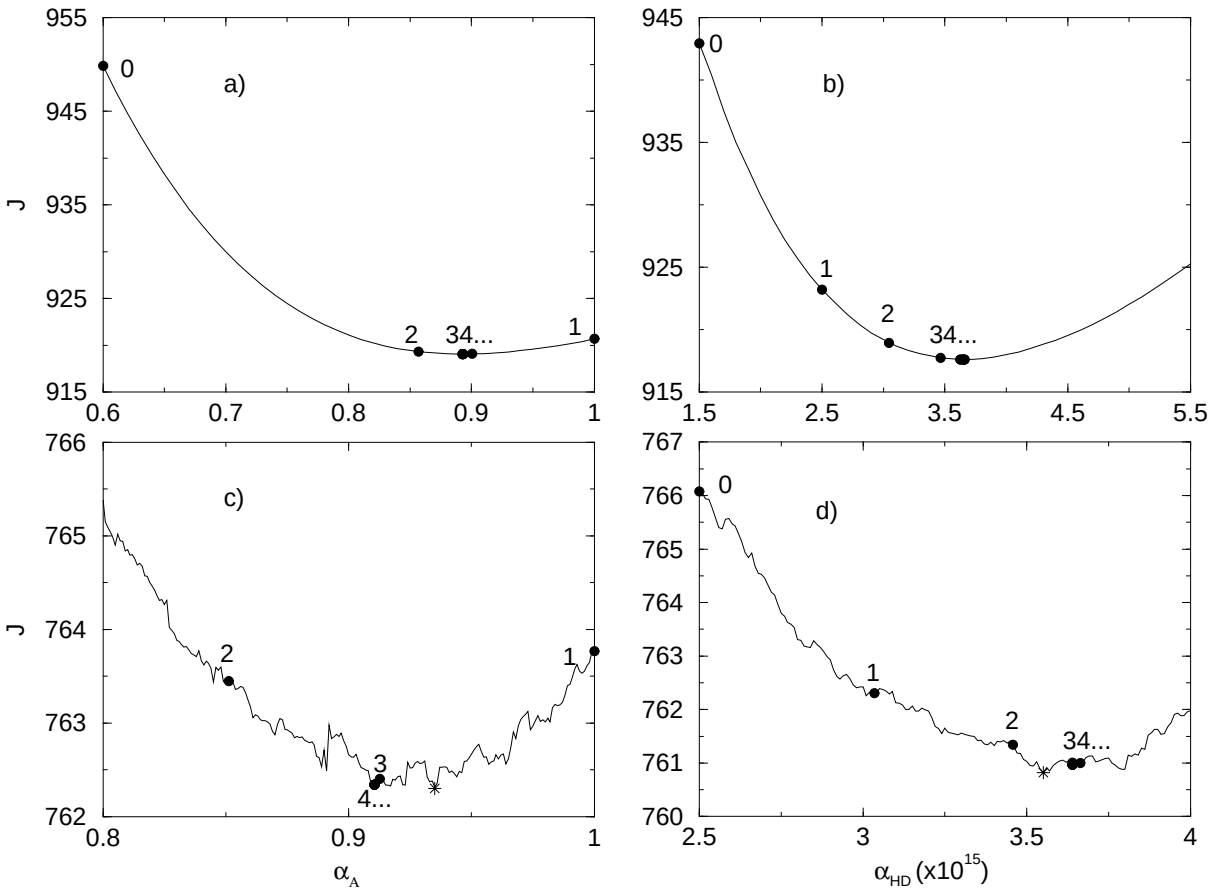

Figure 1: Variations of the cost function (J), defined as a weighted sum of the squared spectral errors of 6-hour forecasts (see (1)), with respect to the time filtering coefficient, α_A (left), and the horizontal diffusion coefficient, α_{HD} (right), using the adiabatic (top) and diabatic (bottom) version of the NCEP model: (a) adiabatic model, α_A, (b) adiabatic model, α_{HD}, (c) diabatic model, α_A and (d) diabatic model, α_{HD}. The cost function is evaluated by a 0.001 interval of α_A and a 0.001×10^{16} interval of α_{HD}. A dark-dot and a number beside it indicate the minimization position and the corresponding iteration number, for solving an optimal parameter value. "Observations" are the NCEP re-analyses at 00 and 06 UTC on 1 October 1995.

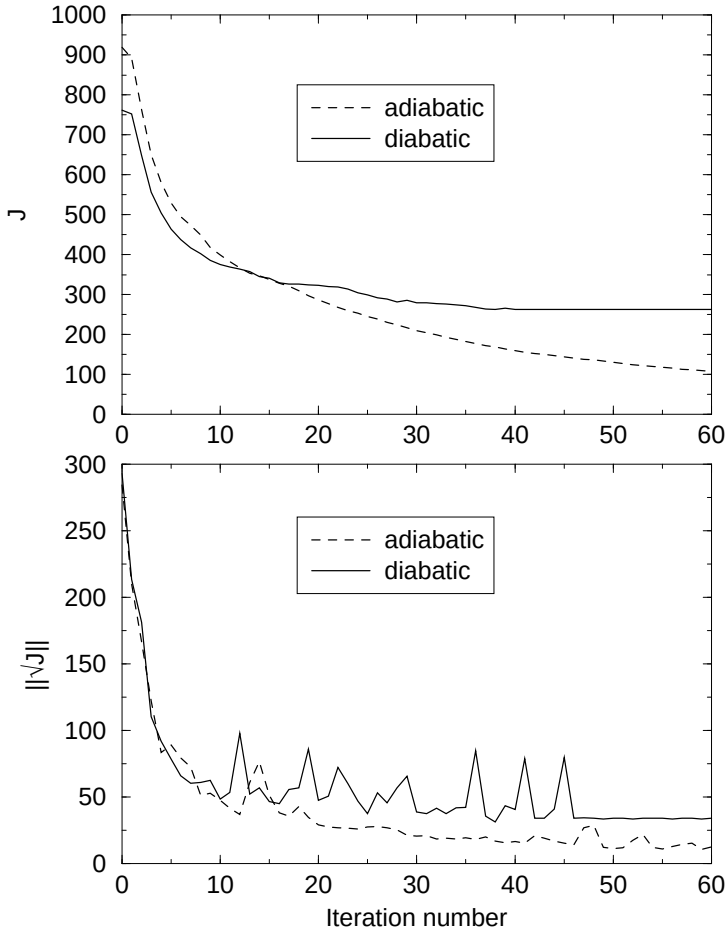

Figure 2: Variations of the cost function (J) (top) and the norm of gradient ($\|\nabla J\|$) (bottom) using the NCEP adiabatic (dashed) and diabatic (solid) global spectral models, with respect to iteration number in the minimization, for optimizing initial conditions.

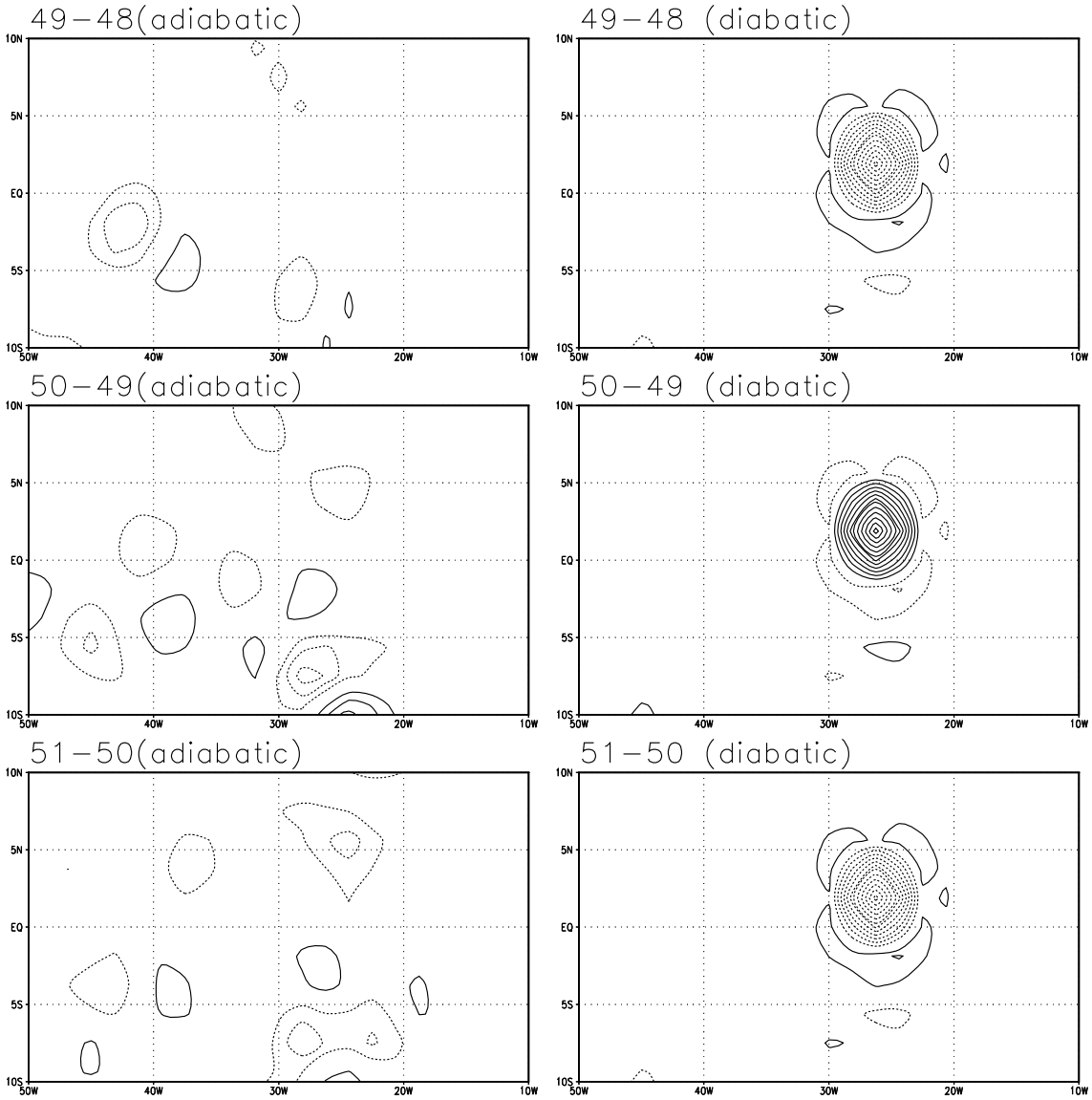

Figure 3: The change of the specific humidity at the 5th model level ($\sigma = 0.931$) at iteration 49 and 48 (top), 50 and 49 (middle), 51 and 50 (bottom) in adaibatic (left) and diabatic (right) 4D-Var minimization. The interval of contours is 0.05×10^{-3} kg kg^{-1} and dashed contours are negative and solid contours are positive.

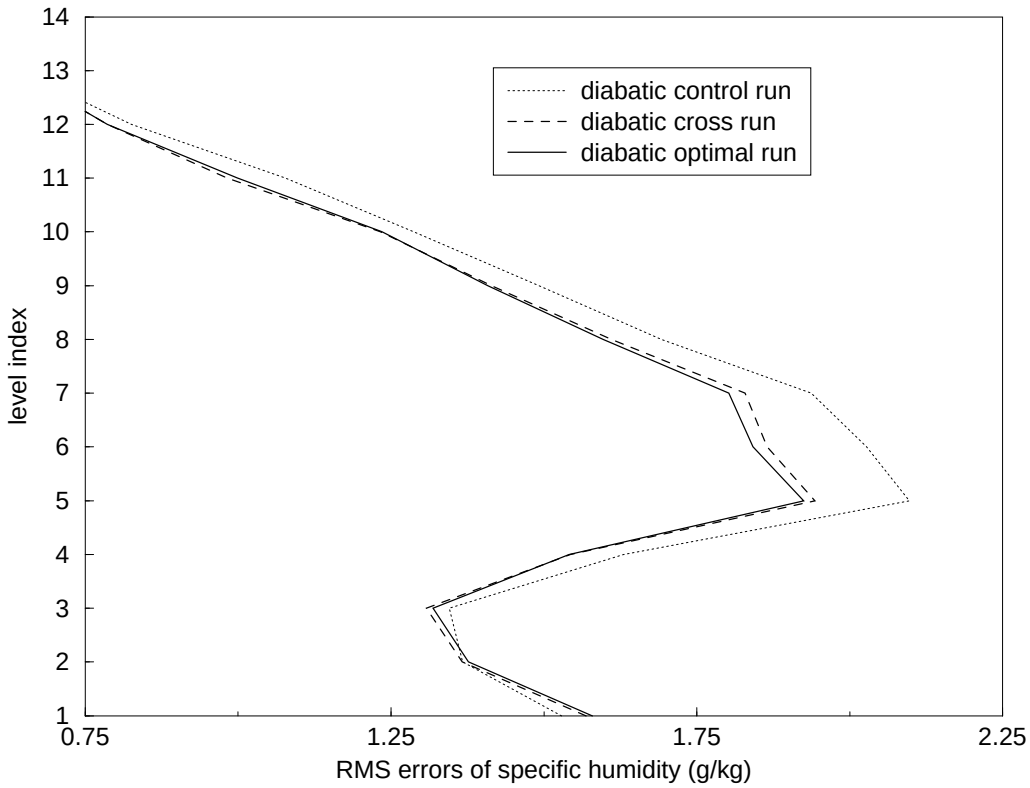

Figure 4: Vertical distributions of RMS errors of specific humidity of 48-hour forecasts over 30°S to 30°N in three diabatic forecast cases: control run (using the NCEP re-analysis as initial conditions: dotted line), diabatic cross run (using the adaibatic optimal initial conditions: thick-dashed line) and diabatic optimal run (using optimal initial conditions solved by a diabatic 4D-Var experiment: thick-solid line) over the global domain.

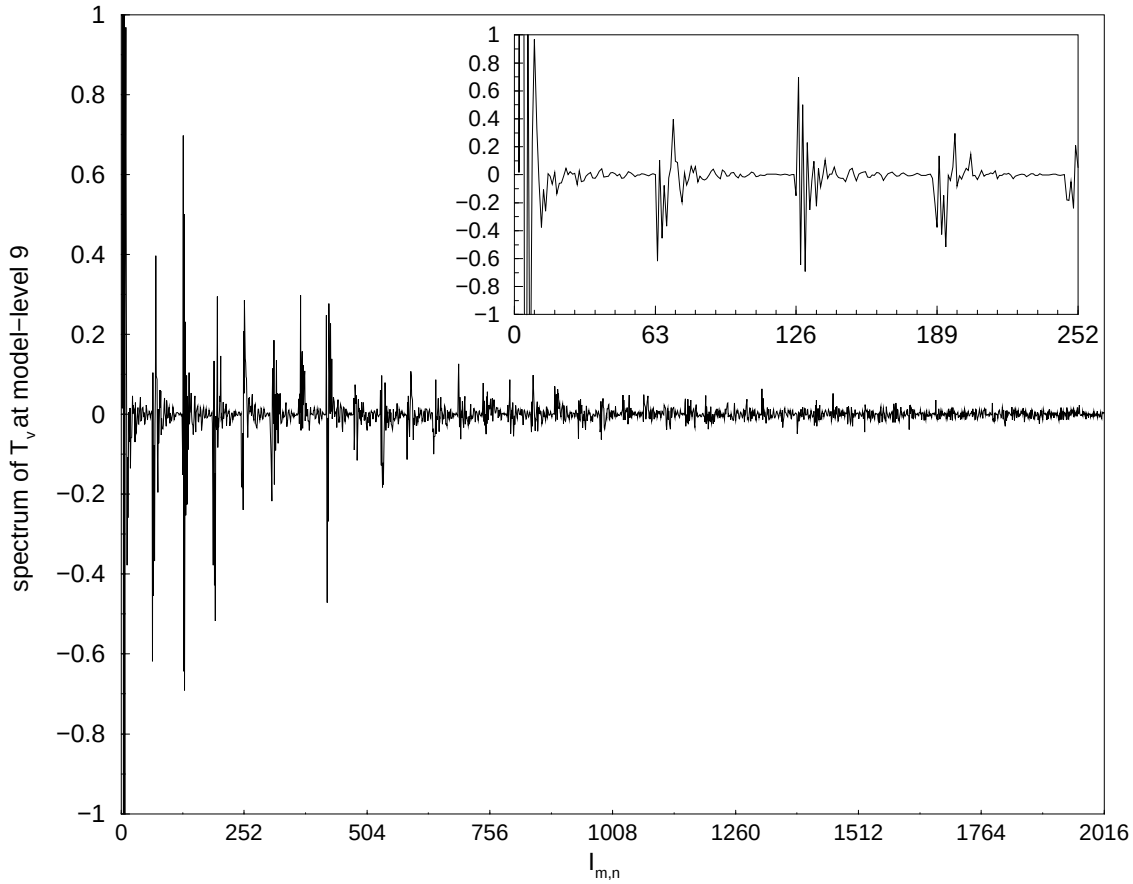

Figure 5: Spectrum of virtual temperature on the 9th model level (from the ground, $\sigma = 0.7508$) with respect to the spectral component index. The first 252 components are magnified in the top-right window.

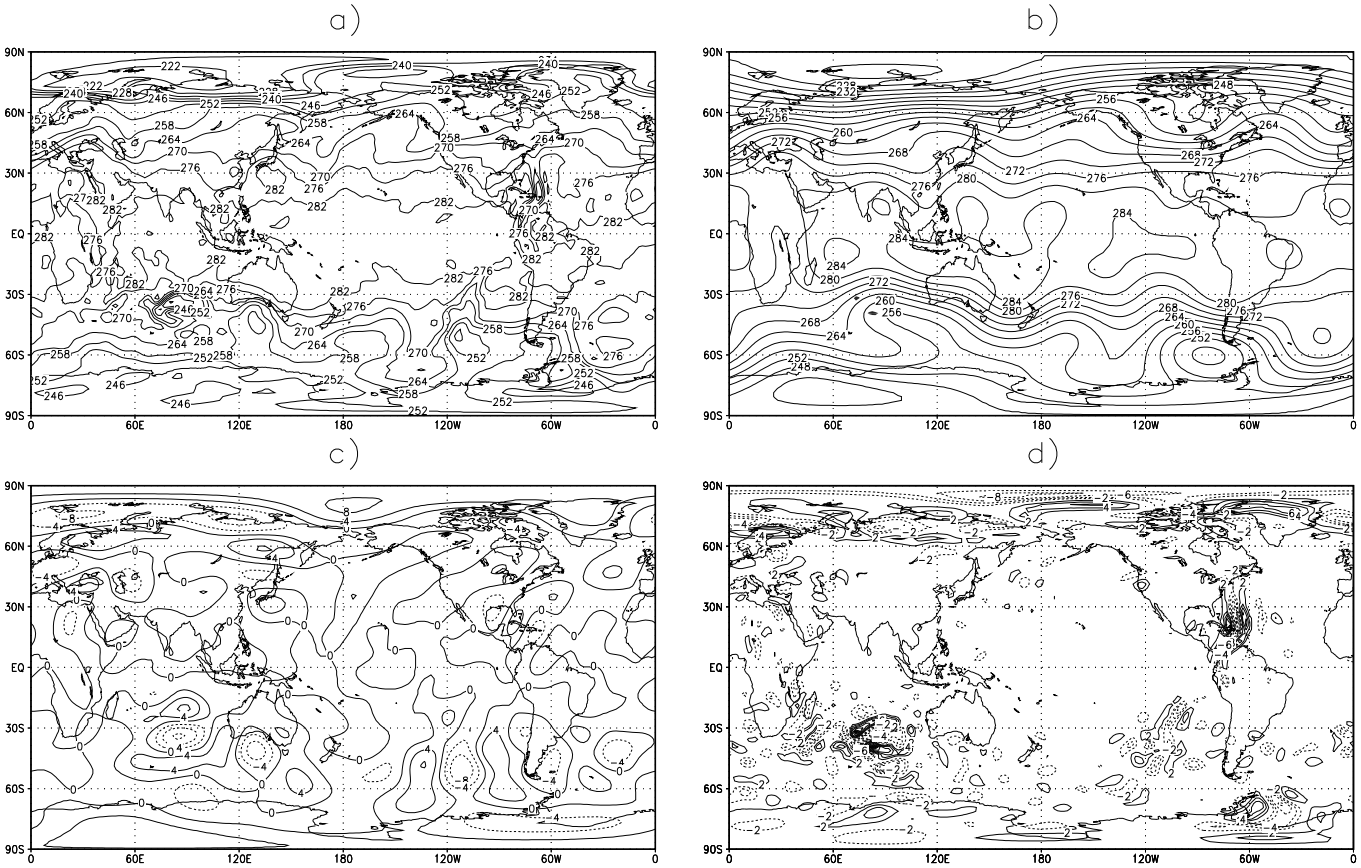

Figure 6: Distributions of a) the original temperature, b) the temperature of "Large-waves" (wavenumbers less than 10), c) the temperature of "Middle-waves" (wavenumbers between 11 and 20) and d) the temperature of "Short-waves" (wavenumbers greater than 20) at the 9th model level, from the NCEP re-analysis at 00 UTC on 1 October 1995. The interval of contours is 4 K for (a), (b) and (c), 2 K for (d).

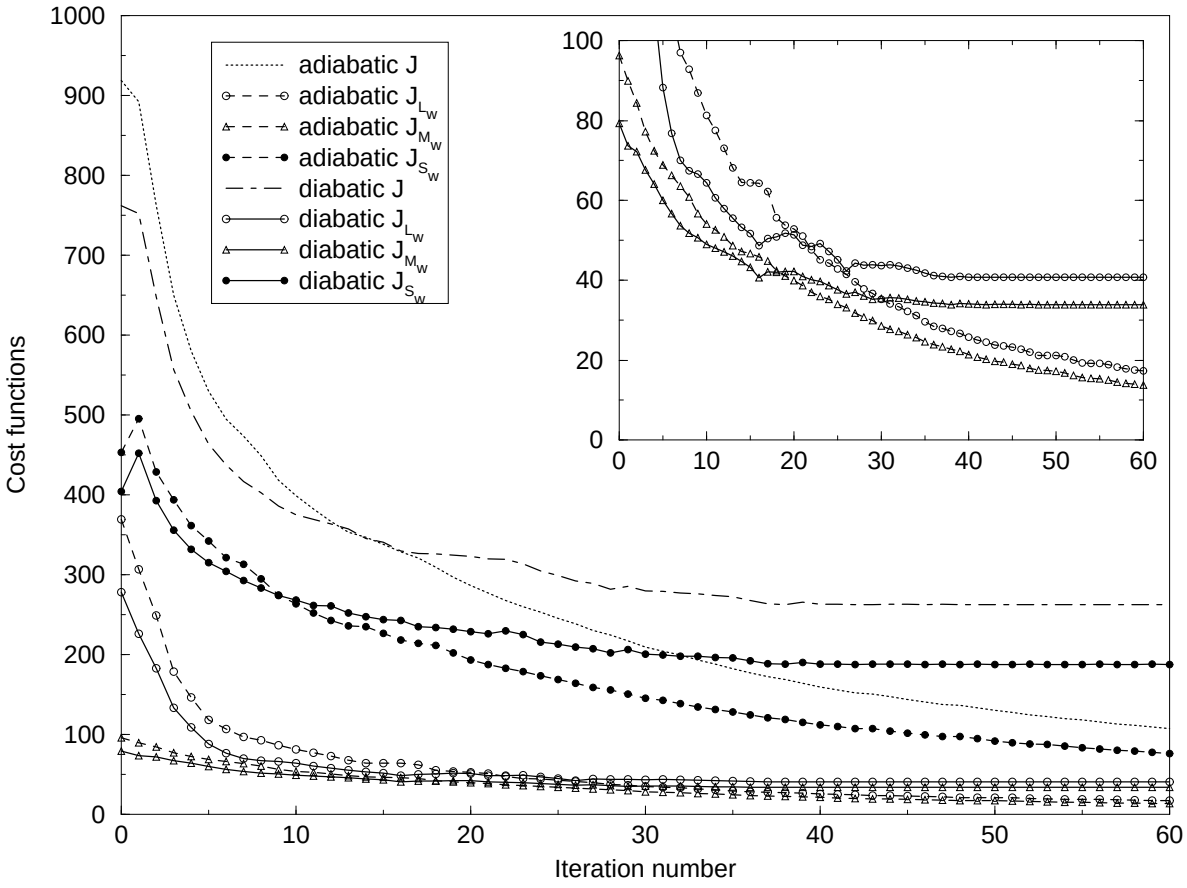

Figure 7: Variations of the cost functions in different wavenumber ranges: "Large-waves" (J_{L_w}), "Middle-waves" (J_{M_w}) and "Short-waves" (J_{S_w}) with iteration number in adiabatic and diabatic cases for optimizing initial conditions. The bottom part (the cost function less than 100) is magnified in the top-right window.

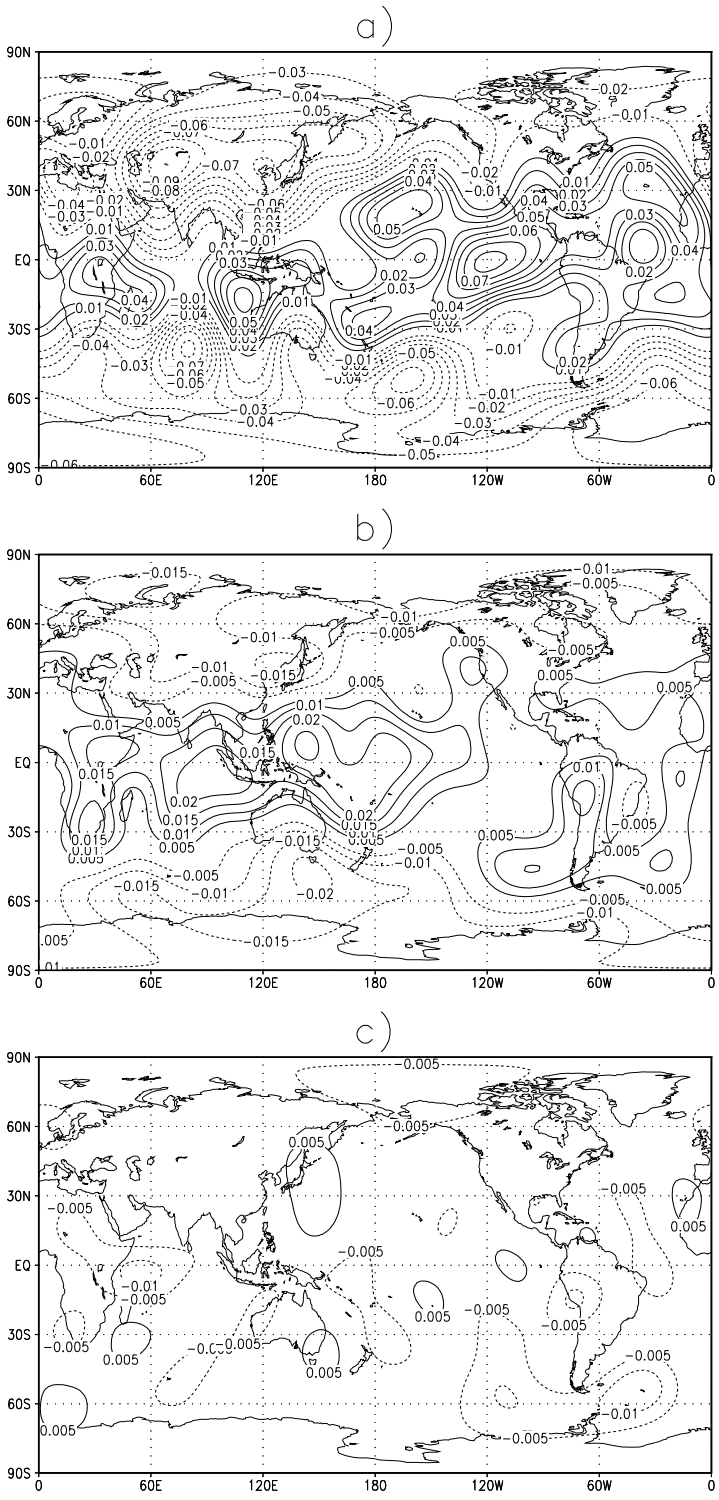

Figure 8: Changes of the "L_w" virtual temperature at the 9th model level ($\sigma = 0.778$) through one-iteration at (a) iteration 5, (b) iteration 20 and (c) iteration 50 in an adiabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c).

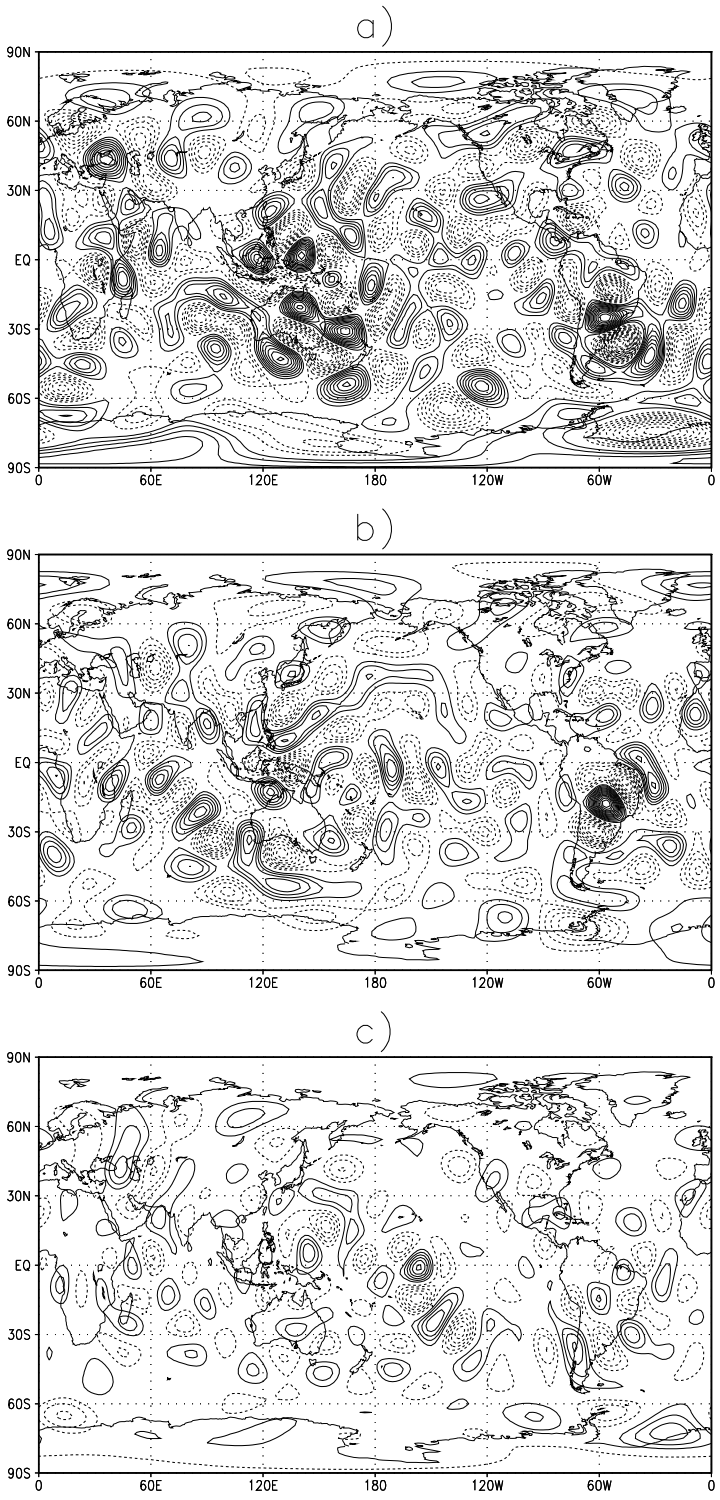

Figure 9: Changes of the "M_w" virtual temperature at the 9th model level ($\sigma = 0.778$) through one-iteration at (a) iteration 5, (b) iteration 20 and (c) iteration 50 in an adiabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c).

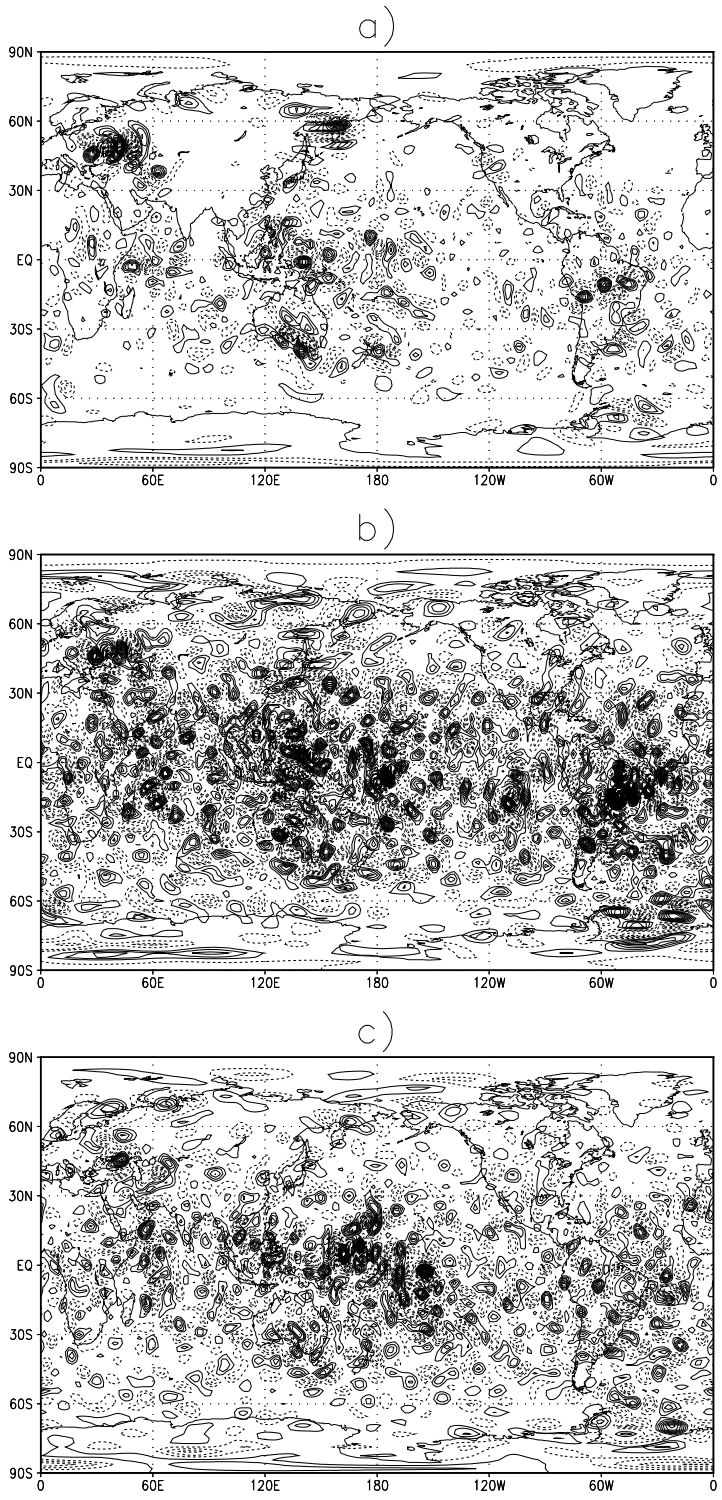

Figure 10: Changes of the "S_w" virtual temperature at the 9th model level ($\sigma = 0.778$) through one-iteration at (a) iteration 5, (b) iteration 20 and (c) iteration 50 in an adiabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c).

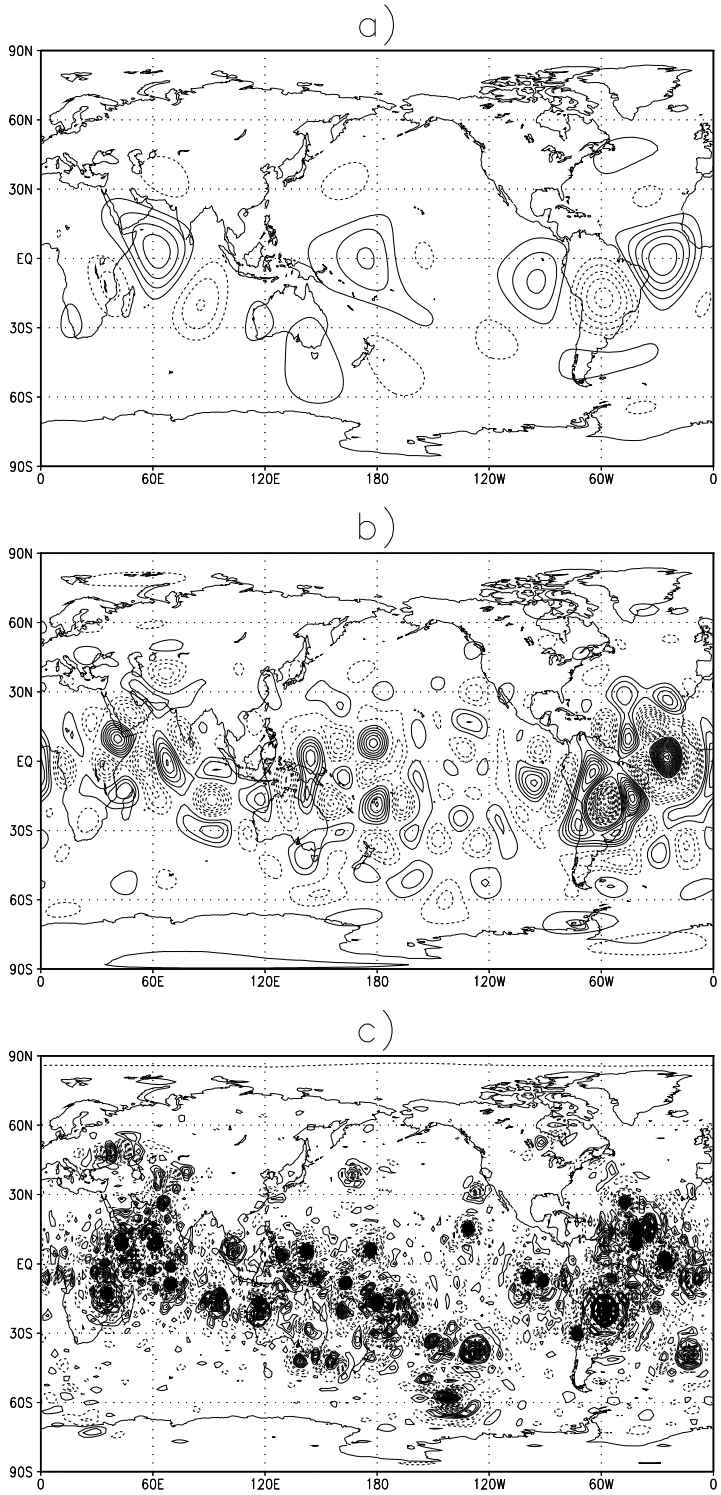

Figure 11: Changes of the (a) "L_w," (b) "M_w" and (c) "S_w" virtual temperature at the 9th model level ($\sigma = 0.778$) through one-iteration at iteration 20 in a diabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c). The dashed contours are negative and solid contours are positive.

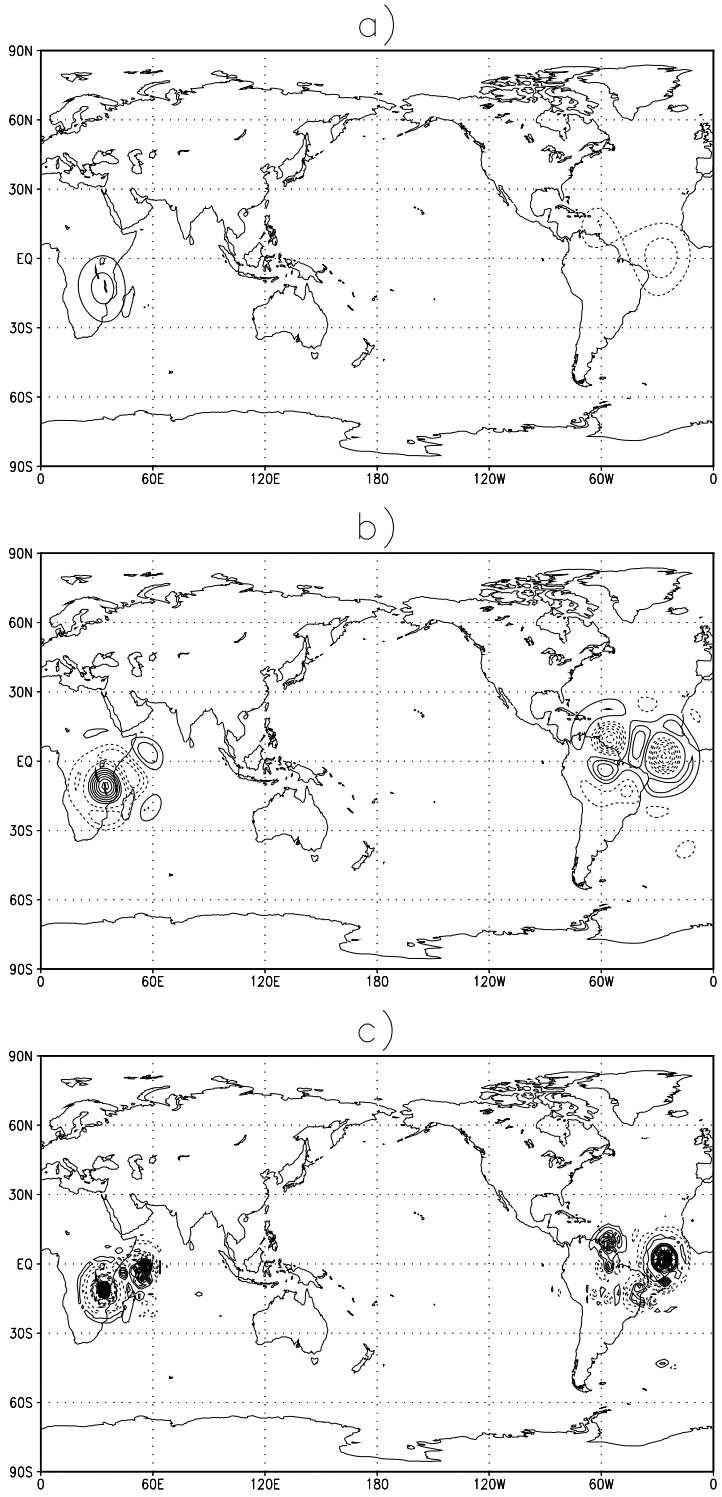

Figure 12: Changes of the (a) "L_w," (b) "M_w" and (c) "S_w" virtual temperature at the 9th model level ($\sigma = 0.778$) through one-iteration at iteration 20 in a diabatic 4D-Var experiment. The interval of contours is 0.01 K for (a) and (b), 0.005 K for (c). The dashed contours are negative and solid contours are positive.

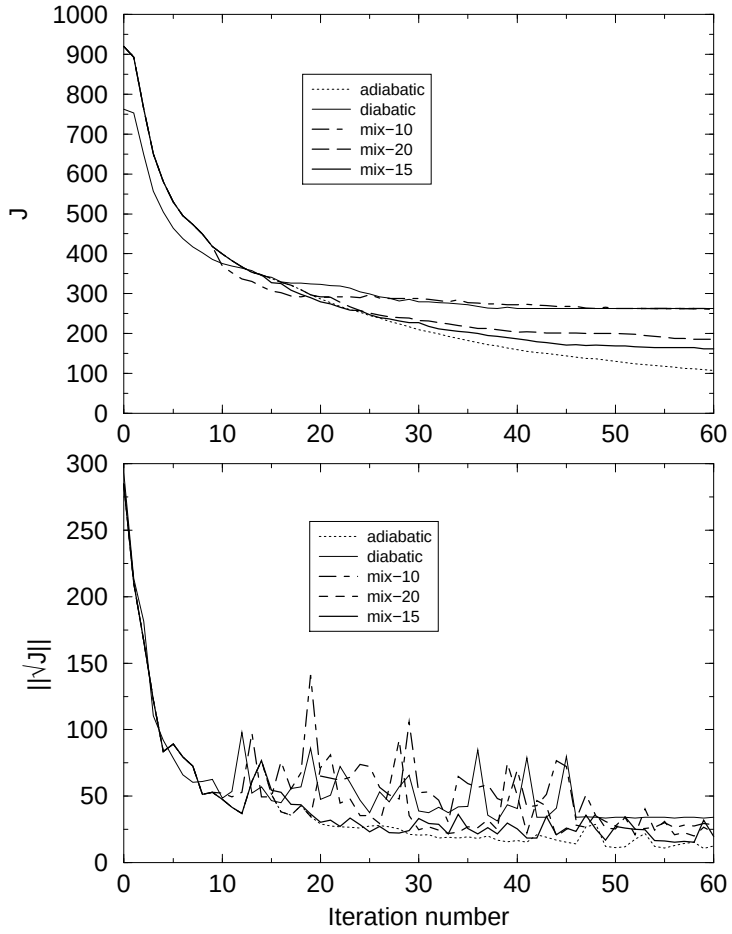

Figure 13: Variations of the cost function (J) (top) and the norm of gradient (bottom) ($\|\nabla J\|$) with iteration number in the new mixed 4D-Var schemes (mix10, mix15 and mix20). The corresponding J and $\|\nabla J\|$ in the old scheme (from Fig. 2) are plotted as reference.

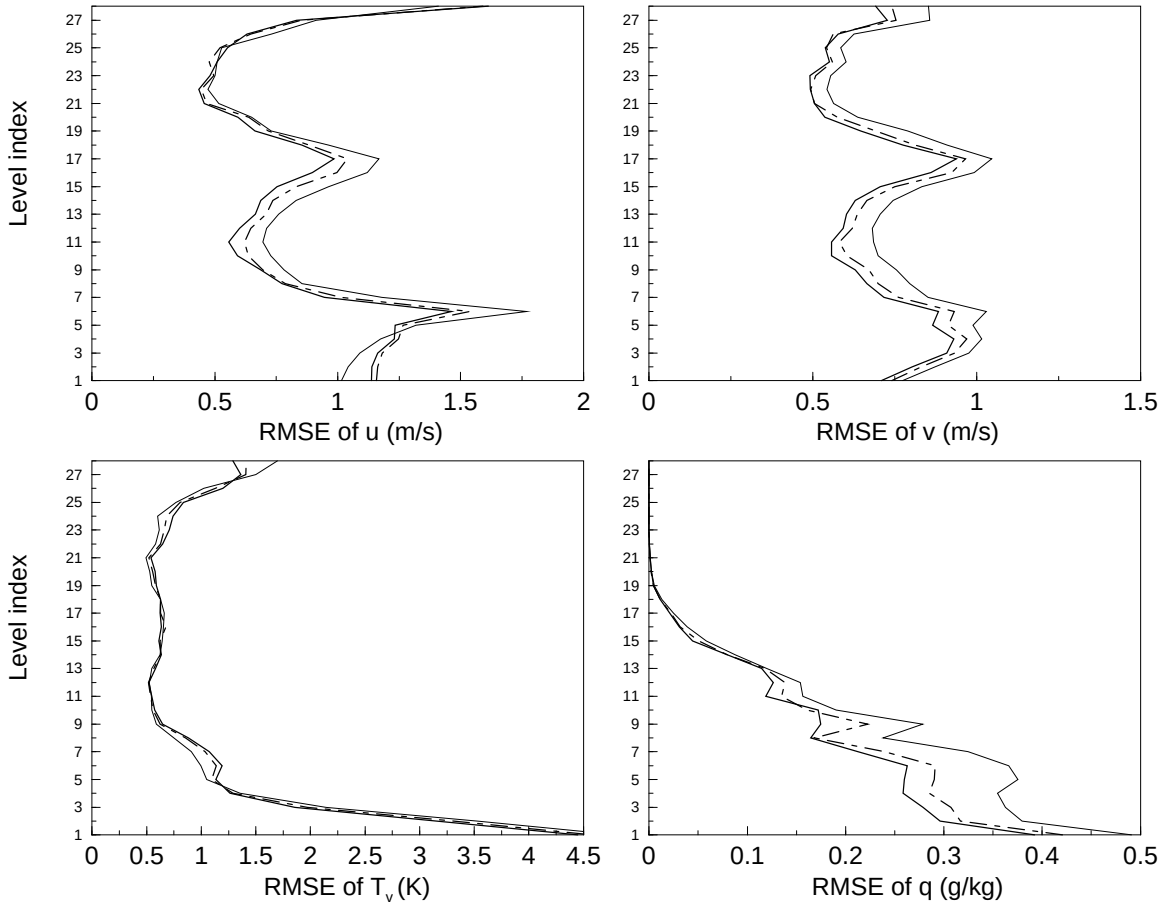

Figure 14: Vertical distributions of the RMS errors of u, v, T_v and q in mix15 (thick-solid), mix20 (dashed) and old (thin-solid) schemes at the end of assimilation window.

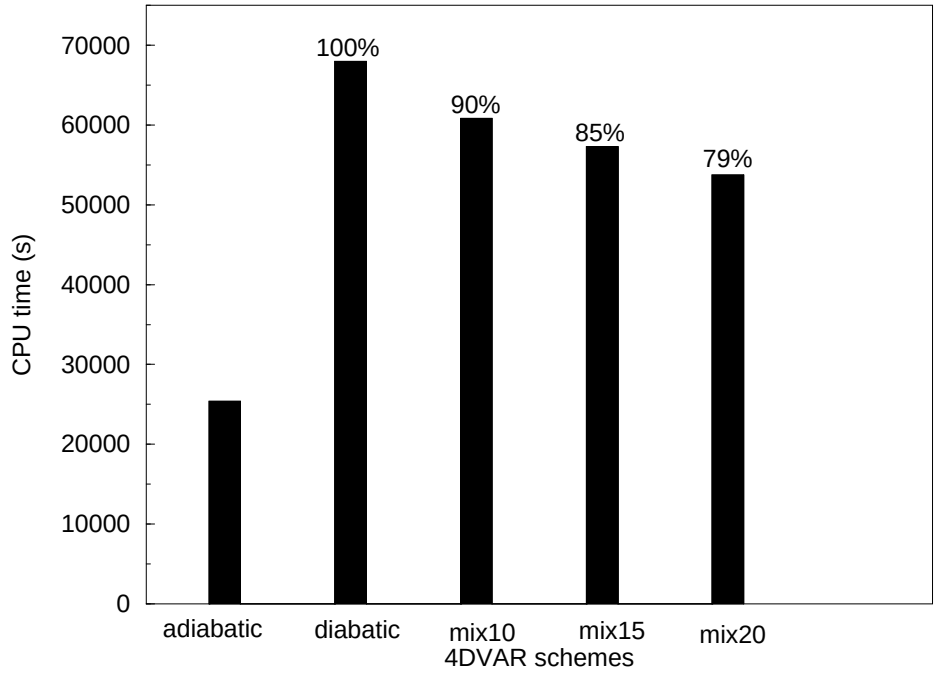

Figure 15: CPU time to conduct 60 iterations for solving the optimal initial conditions in 5 cases.